SUStory: Usability Evaluation of Conversational Interfaces for Children with a Narrative on a Game Board

Preprint · June 2024

DOI: 10.1145/3640794.3665565

3 authors, including:

Shanshan Chen
Eindhoven University of Technology
7 PUBLICATIONS **7** CITATIONS

SEE PROFILE

SUStory: Usability Evaluation of Conversational Interfaces for Children with a Narrative on a Game Board

Shanshan Chen
s.chen1@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Panos Markopoulos
p.markopoulos@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Jun Hu
j.hu@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Figure 1: Traditional SUS and the design of game board-based SUS, SUStory. (A) Traditional SUS for kids; (B) The game board of SUStory; (C) Choice cards of SUStory; (D) Experimental context; (E) SUStory result.

ABSTRACT

The increasing application of conversational agents as assistants and playmates for children brings about the need for evaluation methods tailored for children. The System Usability Scale (SUS) for kids is a well-established instrument for measuring subjective aspects of the usability for children. However, using the scale for children presents difficulties relating to children's concentration and the biases observed when they respond to questionnaires. To make the questionnaire completion enjoyable for children and improve their engagement with the questionnaire, we rendered the adaptation of SUS for children in a game-board format. This paper motivates the presentation of the questionnaire in this way and discusses the use of this instrument in a case study with 35 children aged eight to twelve. We argue that this presentation helped engage with children and there was little evidence of extreme response bias. We conclude that this board format is a more appropriate way to present the SUS questionnaires to children in usability tests, which may also apply to different surveys including rating scales.

KEYWORDS

SUS, usability testing, conversational agents, storytelling, children

ACM Reference Format:
Shanshan Chen, Panos Markopoulos, and Jun Hu. 2024. SUStory: Usability Evaluation of Conversational Interfaces for Children with a Narrative on a Game Board. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3640794.3665565

CUI '24, July 08–10, 2024, Luxembourg, Luxembourg

ACM ISBN 979-8-4007-0511-3/24/07
https://doi.org/10.1145/3640794.3665565

1 INTRODUCTION

The role of conversational agents (CAs) in children's lives is expanding, with applications ranging from language learning support [5] and reading assistance [38] to serving as virtual instructors [32] and enhancing verbal communication skills [7]. As these technologies infiltrate various aspects of daily life—whether through smart home devices like Amazon Alexa and Google Home, or integrated dialogue systems in smartphones like Siri [10][11] – CAs are becoming increasingly popular among children. These agents are present in numerous settings, from solo interactions at home [35] to providing entertainment during family drives [16]. Gradually, CAs are becoming indispensable partners and assistants for children in various forms and situations. Driven by the increasing popularity of CAs engaged in children's daily lives, more summative nature, and quantitative measures of the quality of the interaction such as the usability or the user experience are particularly useful to address the unique challenges of CAs for them involving ethical considerations [23], engagement and usability issues like less precise articulation, limited vocabulary, and fewer strategies for modifying their language [4][11].

Previous research on the usability evaluation methods of CAs targeted at children has focused primarily on two aspects. The first involves behavior observations and analysis of conversational features, such as the intervals of time children spend with agents [7], and audio analysis extracting features like pitch, intensity, and speech rate [27]. These methods, while insightful, often require lab environments or specialized settings that may not be easily accessible to all designers due to relatively high operational or technique costs and limited general applicability.

The second aspect borrows established measurement methods and theories from the field of human-computer interaction (HCI), offering more general and accessible options [31][22]. A notable tool is the System Usability Scale (SUS), favored for its brevity and well-documented psychometric properties. This scale allows comparisons with a vast array of prior tests, providing benchmarks for usability that indicate whether a design is below or above average. The SUS consists of ten items with five Likert response options ranging from strongly disagree to strongly agree. Adaptations, such as those by Putnam et al (2021), have modified the language to ensure comprehensibility and relevance for children [31]. However, there are additional challenges when asking children to complete questionnaires. For example, children lack motivation and may find it tedious to answer questions and find the process of responding to questions tedious. This lack of engagement can lead to non-responses or satisficing behaviors, such as straight-lining—where children might rate all items of the SUS the same without genuinely reflecting on their assessment of the system being evaluated. Additionally, there can be biases associated with extreme responses, where children might strongly agree or disagree with the scale questions [8].

In this paper, we refine Putnam's SUS [31] to advance usability measurement methods for children's CAs, addressing extreme response bias. Our research, stimulated by a qualitative comparison of a sleep diary chatbot with text and voice interfaces, revealed limitations in the traditional SUS. We redesigned the SUS into a game board format with narrative elements to better suit young children. Our goal is to establish a reliable and enjoyable method for evaluating CAs for children, highlighting the need for innovative approaches in assessing conversational interfaces and child-centric applications.

2 PRIOR WORK

2.1 Related Work

Children frequently rate scales at the extremes, which can lead to misleading interpretations of their responses [8][18][39]. Unique traits such as concentration abilities, capacity for abstract thinking, adaptability, and goal monitoring significantly influence their participation in usability testing [25]. To counteract extreme response bias, it is crucial to consider these characteristics when designing usability questionnaires for children.

The power of storytelling lies in its ability to immerse users within a narrative context, effectively bridging the gap between their experience and measurement tools [33]. Designers often leverage the immersive nature of storytelling to enhance usability testing [20], aiding in the development of high fidelity and their preparation for market launch as products or services [12][13]. Storytelling proves especially beneficial in children's application design, significantly increasing engagement [29][2]. Consequently, we assert that integrating storytelling can significantly improve usability test for children, particularly by enhancing their concentration.

Physical engagement can enhance cognitive processes, particularly when combined with narrative immersion, helping children to better retain and understand information [8]. This effect has been proved in a user test case involving tabletop and room-based games [37]. Gamified experiences further amplify these effects by keeping children focused on their tasks and providing clear markers of their progress toward goals. Based on these insights, we propose that conducting usability tests in a physical game or game-like format could effectively mitigate the bias caused by a lack of engagement and concentration.

2.2 Pilot Study

In our pilot study aimed at developing a sleep diary chatbot for children, we compared a text-based chatbot (Snoozy) [1] with a voice-based chatbot (Dozzz) [9]. We involved five children (M = 9.8 years old, SD = 1.30) in a within-subject experiment where each child interacted with both chatbot versions and then completed the SUS for kids questionnaire (Figure 1 (A)) [31] for each. To minimize order and practice effects, the procedure was counterbalanced, with three children starting with the voice version and two with the text version. Behavior was recorded via camera.

Despite industry SUS score averages of 68-70.5, both interfaces in our study scored above 90, indicating no significant difference and suggesting extreme response bias. This outcome suggests that the SUS for kids appeared to suffer from extreme response bias, as it failed to differentiate between the two interfaces – our primary objective. This concern was supported by the interview data, where children expressed a preference for the voice-based interface, something that the SUS scores did not manage to capture. Additionally, analysis showed SUS scores among children mostly clustered at the extremes, either "1" (Strongly Disagree) or "5" (Strongly Agree), with few moderate scores. This pattern points to potential response biases or limitations in the scoring system to capture nuanced experiences, as shown in Figure 2 for one of the chatbots.

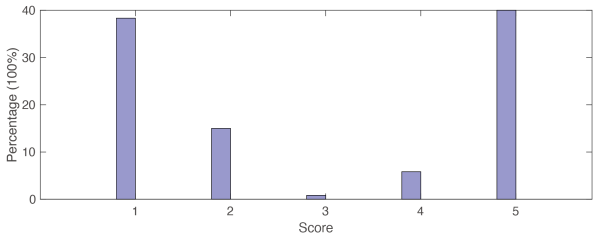

Figure 2: The percentage of each score in SUS for kids

Ideally, respondents in evaluation are expected to provide 'optimizing' responses, which naturally vary [36]. However, our pilot study revealed a troubling homogeneity in extreme responses in SUS for kids, raising concerns about potential data contamination linked to issues with concentration. Notably, 65% of items were answered in less than four seconds, which is below the threshold value of cognitive speed expected for children around ten years old [1][6][19]. This rapid response rate raises concern about their concentration during the experiment and raises suspicion for satisficing in answering the survey items [26]. Consequently, we hypothesize

[1]Adults typically read at a speed of 250 to 300 words per minute[6]. When reading an 11-word sentence (the average length of words per question in SUS for kids) it should take an adult about 1.5 to 3 seconds. However, children around ten read approximately 1.8 times slower than adults[19]. Consequently, the time threshold for children to read an 11-word sentence is estimated at about 4 seconds.

that combining storytelling with interactive game elements can create an engaging environment that enhances children's focus and participation in testing.

3 DESIGN OF SUSTORY

Building the above hypothesis, we developed SUStory, a game-board adaptation of the SUS for children aged eight to twelve, enhancing engagement through storytelling. Using the character Dozzz from our voice-based chatbot as an example, we structured its development into three steps to ensure an engaging experience for children.

Craft a Backstory for the Agent. Since the character of an agent can be animals or virtual humans, the backstory starts with the character's progress and personal development by overcoming difficulties that match the questions in the questionnaire. The plot can come from famous fables or general stories, like finding friends by completing questions. For example, Dozzz, the avatar in our voice-based chatbot, is a duck [9]. Our storytelling journey began with Andersen's "*The Ugly Duckling*" to create SUStory: "*Finding Dozzz.*" In this adventure, Dozzz, a duckling dreaming of becoming a swan, faces challenges such as neighbourly mockery and harsh weather. As they answer questions from the SUS for kids, the story progresses, culminating in Dozzz's transformation into a swan.

Enrich the Story to Align with the Questionnaire. To better integrate the character's backstory with the questionnaire, we enriched the story with additional challenges. These challenges are portrayed as steps/challenges that respondents overcome by answering questions. In Dozzz's story, we created a colorful and whimsical painting that shows his journey from a duck to a swan. The SUS for kids [31] questions are embedded in clouds, representing the challenges Dozzz faces. As children answer each question, they progress through these clouds, mirroring Dozzz's adventure. The visual journey begins with Dozzz's birth on a tree, moving through scenes of sharing dreams with neighbors and adventures across landscapes, ending with his transformation at the top of a tree (as shown in Figure 1 (B)).

Play in SUStory. To enhance engagement, we integrated the storytelling element of SUStory into a game board format, leveraging the benefits of board games in encouraging communication and enjoyment among children [34]. Instead of marking selections on paper (Figure 1 (A)), we've eliminated the color from five options (Figure 3(A)) and created colorful icon cut-out cards (Figure 3 (B)). Children can then place these vibrant cards on their chosen selections, effectively indicating their answers (Figure 3 (C)).

4 EVALUATION

We set out to evaluate how well SUSstory serves as a way of presenting the SUS questionnaire. 35 children aged eight to twelve used SUStory to evaluate a chatbot interface. We checked for extreme response bias which is often the case when children fill in evaluation questionnaires and we conducted interviews to understand their experiences with SUStory.

4.1 Participants

We recruited 35 children, aged eight to twelve years (M = 9.3, SD = 1.19), with parental consent and voluntary, anonymous participation. Ethical approval was obtained from our university's Ethical Review Board, and the experiment was conducted at the university premises in November 2023. The duration was limited to 20 minutes, and behavior was recorded via camera.

4.2 Procedure

Following an initial introduction and familiarization with the chatbot, children began the test session. They interacted with Dozzz, then completed the SUStory questionnaire. Subsequently, we conducted semi-structured interviews to gather feedback on their experience with SUStory. We will now detail the process of the usability test with SUStory:

The session began with the experimenter leading the participant to the game board and introducing the story: *"You are now Dozzz, a character from the voice-based sleep diary. You are a duck dreaming of becoming a swan to soar through the skies, you'll face challenges. I'll be here to help you conquer these obstacles and achieve your dream. Are you ready to embark on this thrilling adventure with me?"* To this, all children responded enthusiastically. The experimenter continued, *"You're a baby in your tree home, telling friends and neighbors about your dream. They doubt you. But your task is to filter out the negativity and stay true to your dream by answering the first question here."*, pointing to the first cloud on the board. After addressing this query, the experimenter narrated the next part of the journey: *"As you leave your neighborhood, you must cross a river, answered by the next question."* This immersive process continued, leading children through different story parts as they chose responses to each query independently (Figure 4).

Acting as a storyteller, the experimenter immersed participants in the world of Dozzz, seamlessly narrating until the final challenge was overcome, culminating in congratulations for aiding Dozzz in achieving his dream. Most sessions were completed within three minutes, offering rich insights into each child's engagement and individual thinking, as shown in Figure 1 (D). Importantly, the story just bridged the gap between successive questions, prompting children to proceed to the next one.

4.3 Results

We monitored the responding speed of children while they answered questions in SUStory. While only 11.6% of the children responded to the items in less than the threshold value of four seconds [6][19]. However, a one-sample t-test showed no significant difference between the sample mean (4.10) and the threshold of 4.0 ($t(34) = 0.7391, p = 0.4649$). The confidence interval included the threshold value, supporting this conclusion. We do not have sufficient samples from the pilot study for a similar t-test, but the average response time there is 2.06s ($SD = 0.64$). These results suggest that SUStory aligns better with the threshold value. Further analysis of extreme responses will be conducted to understand SUStory's impact.

Evaluation of Extreme Response Bias. For assessing extreme response bias, we deployed the Representative Indicator for Response

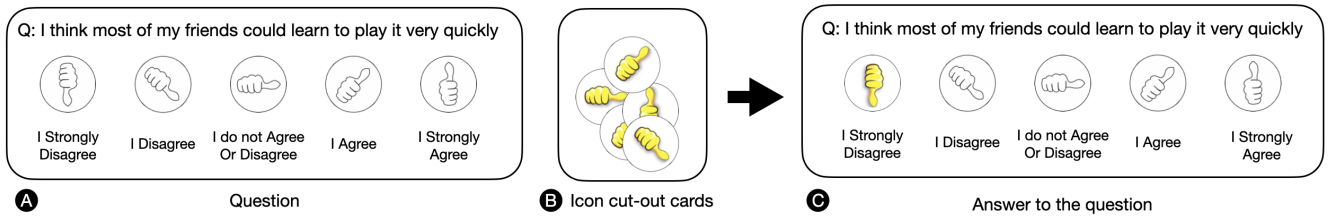

Figure 3: The question in SUStory

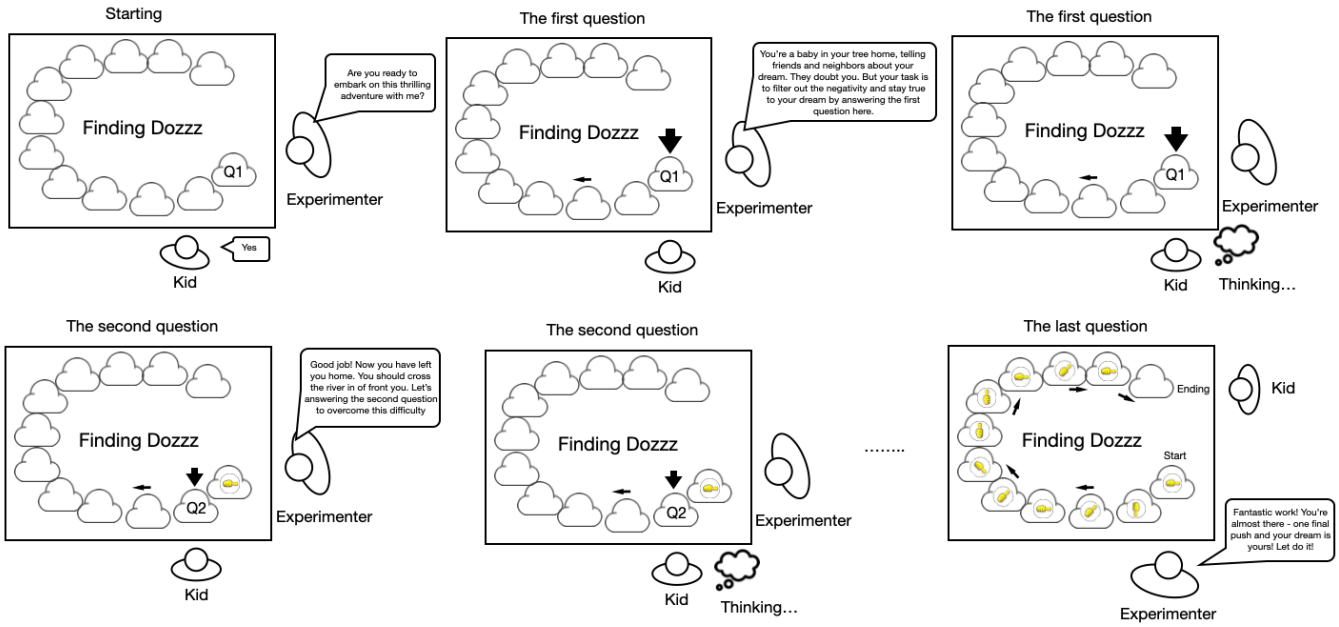

Figure 4: The experiment procedure

Style, RIRS [17], recommended by many researchers duo to its effectiveness in avoiding the effects of homogeneity questions [3]. Our primary focus was on extreme response style (ERS) by dividing the number of extreme responses in the heterogeneity by the total number of items. The overall ERS for the SUStory was low ($M = 0.1389, SD = 0.1086$), which suggests that the extreme response tendencies noted in the pilot has been successfully mitigated. While the size of the pilot does not support statistical comparison, we note that the ERS in the limited data of the pilot study was higher ($M = 0.2, SD = 0.095$).

Differentiation of Responding. We display the percentage of each score (from 1 to 5) for Dozzz. Figure 5 illustrates how children provided scores within the middle range for SUStory. This observation suggests a differentiation of ratings among participants, indicating a nuanced and varied response pattern.

Perception. Almost all of the children expressed their appreciation of the board game format of the questionnaire. One noted, *"I like the colors and characters. I can help the character. I enjoy playing with them and learning alongside them."* (P27). Others appreciated listening to the story's narrative and interacting on the board rather than selecting answers on paper, which felt like an exam (P24, P30).

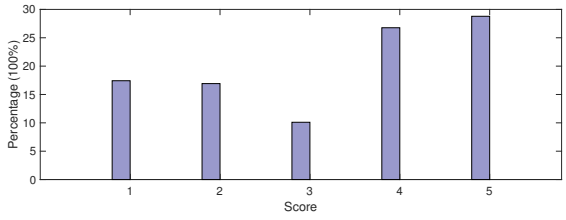

Figure 5: The proportion of each score

5 DISCUSSION

Our evaluation found that children actively engaged with SUStory, enhancing their playful experience without exacerbating biases typical in traditional questionnaires. This approach, integrating a chatbot avatar with a storyline, is broadly applicable in usability evaluations for children's CAs. It offers enjoyment for participants and yields reliable results for designers without imposing significant burdens.

5.1 Supporting concentration through physical engagement

Concentration is vital in children's usability tests [25], influencing their participation and feedback quality. Literature suggests that physical engagement, combined with narrative immersion, enhances cognitive processes and information retention [15]. Therefore, incorporating physical activities could significantly boost concentration during usability testing for children.

Using a game board format where children indicate preferences through physical actions significantly enhanced engagement. Interacting with the board and icon cards—like placing cards or moving around the board—deepened their immersion in the narrative. This active participation was evident in behaviors such as positioning icons and consistent thumb directions in response to questions, as shown in one child's result (Figure 1 (E)).

Moreover, the design of SUStory, resembling a game board and incorporating captivating visuals and vibrant color schemes also enhances the physical engagement by creating an interactive and enjoyable environment. This approach is in line with the established benefits of traditional board games, which are known for fostering open-mindedness and contributing to the enhancement of literacy [24], cognitive skills [14], and aesthetics [30].

Research suggests that engaged children tend to spend more time on tasks [6][19]. Our observations confirm this, demonstrating a significant reduction in the proportion of questions answered under the process speed threshold with the board-format questionnaire compared to the traditional A4 paper-based method used in our pilot study. This indicates the potential for enhanced concentration on activities when using SUStory. The children preferred to view each question as an opportunity to help their avatar advance toward a goal, rather than simply marking answers on paper. This behavior highlights the suitability of physical activities on the board for enhancing children's concentration and engagement with the material.

5.2 Game board-format promote monitoring progress

The ability to monitor progress toward a goal is a pivotal factor in children's usability testing [25], linked to the processes of outcome evaluation and redirection of unsuccessful efforts [21]. While children might not naturally excel in these skills [21], the use of a game board can effectively support their development because games provide clear challenges and targets, along with a sense of mastery [28], which are instrumental in promoting progress monitoring. In our study, the game board format was employed to clearly delineate targets and visually appealingly present them, thereby facilitating progress monitoring for children. This approach made it straightforward for children to track their progress, with the experimenter's narrative further enhancing their understanding. At crucial junctures, the experimenters indicated proximity to the end of the narrative conclusion and encouraged them to persevere. This strategy differs significantly from the traditional SUS approach, which relies heavily on self-regulation[21].

5.3 Limitations and Future Work

Our research has several limitations that warrant acknowledgment. The small sample size of our pilot study limits our ability to make full comparisons with the results from SUStory. Future studies with larger sample sizes are recommended to validate these findings and further explore the differences between the SUStory and traditional group. Additionally, we did not examine performance differences or acceptability across various age groups, leaving the effectiveness of this approach for diverse ages unexplored. Further research with a larger, robust sample is crucial to confirm the reliability and validity of SUStory, particularly across different age groups. Secondly, while our version of SUS for kids offers insights within the context of designing conversational interfaces for sleep diaries, its broader application, such as in educational settings, needs exploration. Future studies should also evaluate how questionnaires are presented to children, identifying potential biases and determining at what age such methods may be perceived as too childish compared to traditional text-based forms.

6 CONCLUSIONS

In this study, we introduced SUStory, a game-board-based adaptation of the SUS questionnaire for children aged eight to twelve. The board-game structures the task of filling in a questionnaire and provides a sense of progress and closure naturally. We have demonstrated how SUStory was applied to assess the subjective usability of two chatbot interfaces for this age group. On the positive side, children were engaged with filling in the questionnaire and considered all items before responding. While they were relatively slow in answering the questionnaire items, we found no evidence of extreme response bias or satisficing which typically occurs when children use rating scales in usability testing. Children were positive about using this instrument, which suggests that it helps reduce the burden of self-report in usability testing. Future investigations with SUSstory could seek evidence on the reliability and validity of responses, especially examining potential framing effects. Furthermore, it could examine how this approach should be better targeted to narrow target age groups.

REFERENCES

[1] Tessa Aarts, Panos Markopoulos, Lars Giling, Tudor Vacaretu, and Sigrid Pillen. 2022. Snoozy: A chatbot-based sleep diary for children aged eight to twelve. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference*. 297–307.

[2] Agnese Addone, Renato De Donato, Giuseppina Palmieri, Maria Angela Pellegrino, Andrea Petta, Vittorio Scarano, and Luigi Serra. 2021. Novelette, a usable visual storytelling digital learning environment. *IEEE Access* 9 (2021), 168850–168868.

[3] Hans Baumgartner and Jan-Benedict EM Steenkamp. 2001. Response styles in marketing research: A cross-national investigation. *Journal of marketing research* 38, 2 (2001), 143–156.

[4] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[5] Neelma Bhatti, Timothy L Stelter, Scott McCrickard, and Aisling Kelliher. 2021. Conversational user interfaces as assistive interlocutors for young children's bilingual language acquisition. In *Proceedings of the 2021 ACM International Conference on Interactive Media Experiences*. 208–211.

[6] Marc Brysbaert. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of memory and language* 109 (2019), 104047.

[7] Fabio Catania, Micol Spitale, Giulia Cosentino, and Franca Garzotto. 2021. Conversational agents to promote children's verbal communication skills. In *Chatbot Research and Design: 4th International Workshop, CONVERSATIONS 2020, Virtual Event, November 23–24, 2020, Revised Selected Papers 4*. Springer, 158–172.

[8] Christine T Chambers and Charlotte Johnston. 2002. Developmental differences in children's use of rating scales. *Journal of pediatric psychology* 27, 1 (2002), 27–36.

[9] Shanshan Chen, Panos Markopoulos, and Jun Hu. 2023. Dozzz: Exploring Voice-based Sleep Experience Sampling for Children. In *Proceedings of 17th EAI International Conference on Pervasive Computing Technologies for Healthcare*.

[10] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with computers* 31, 4 (2019), 349–371.

[11] Radhika Garg and Subhasree Sengupta. 2020. He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.

[12] Guiseppe Getto. 2020. The story/test/story method: A combined approach to usability testing and contextual inquiry. *Computers and Composition* 55 (2020), 102548.

[13] Guiseppe Getto and Christina Moore. 2017. Mapping personas: Designing UX relationships for an online coastal atlas. *Computers and Composition* 43 (2017), 15–34.

[14] Jeffrey Goldstein. 2012. *Play in children's development, health and well-being*. Toy Industries of Europe Brussels.

[15] Thalia R Goldstein, Brittany N Thompson, and Pallavi Kanumuru. 2022. Do embodiment and fictionality affect young children's learning? *Journal of Experimental Child Psychology* 213 (2022), 105275.

[16] Michal Gordon and Cynthia Breazeal. 2015. Designing a virtual assistant for in-car child entertainment. In *Proceedings of the 14th International Conference on Interaction Design and Children*. 359–362.

[17] Eric A Greenleaf. 1992. Measuring extreme response style. *Public Opinion Quarterly* 56, 3 (1992), 328–351.

[18] Jean-Xavier Guinard. 2000. Sensory and consumer testing with children. *Trends in Food Science & Technology* 11, 8 (2000), 273–283.

[19] Sandra Hale. 1990. A global developmental trend in cognitive processing speed. *Child development* 61, 3 (1990), 653–663.

[20] Tina M Hickey and Anne-Marie de Mejia. 2014. Immersion education in the early years: A special issue. , 131–143 pages.

[21] Berk Laura. 2018. *Exploring Child & Adolescent Development*. Pearson.

[22] Jingya Li, Tengjia Zuo, Erik Van Der Spek, and Jun Hu. 2024. Animated Scale: Adaption of the Motivational Scale for User Testing with Children. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[23] Silvia B Lovato, Anne Marie Piper, and Ellen A Wartella. 2019. Hey Google, do unicorns exist? Conversational agents as a path to answers to children's questions. In *Proceedings of the 18th ACM international conference on interaction design and children*. 301–313.

[24] Karen Markey, Fritz Swanson, Andrea Jenkins, Brian J Jennings, Beth St. Jean, Victor Rosenberg, Xingxing Yao, and Robert L Frost. 2008. Designing and testing a web-based board game for teaching information literacy skills and concepts. *Library Hi Tech* 26, 4 (2008), 663–681.

[25] Panos Markopoulos and Mathilde Bekker. 2003. On the assessment of usability testing methods for children. *Interacting with computers* 15, 2 (2003), 227–243.

[26] Panos Markopoulos, Janet C Read, Stuart MacFarlane, and Johanna Hoysniemi. 2008. *Evaluating children's interactive products: principles and practices for interaction designers*. Elsevier.

[27] Ivonne Monarca, Franceli L Cibrian, Angel Mendoza, Gillian Hayes, and Monica Tentori. 2020. Why doesn't the conversational agent understand me? a language analysis of children speech. In *Adjunct proceedings of the 2020 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2020 ACM international symposium on wearable computers*. 90–93.

[28] Cheryl K Olson. 2010. Children's motivations for video game play in the context of normal development. *Review of general Psychology* 14, 2 (2010), 180–187.

[29] Jaehyun Park, Noha A Mostafa, and Hyo-Joo Han. 2020. "StoryWeb": A storytelling-based knowledge-sharing application among multiple stakeholders. *Creativity and Innovation Management* 29, 2 (2020), 224–236.

[30] Celia Pearce. 2006. Games as art: The aesthetics of play. *Visible Language* 40, 1 (2006), 66.

[31] Cynthia Putnam, Melisa Puthenmadom, Marjorie Ann Cuerdo, Wanshu Wang, and Nathaniel Paul. 2020. Adaptation of the system usability scale for user testing with children. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–7.

[32] Diana Pérez-Marín and Ismael Pascual-Nieto. 2013. An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology* 32, 9 (2013), 955–964.

[33] Christian Roth and Hartmut Koenitz. 2016. Evaluating the user experience of interactive digital narrative. (2016), 31–36.

[34] Heather Smith. 2006. Playing to learn: A qualitative analysis of bilingual pupil-pupil talk during board game play. *Language and Education* 20, 5 (2006), 415–437.

[35] Jennifer Ureta, Celina Iris Brito, Jilyan Bianca Dy, Kyle-Althea Santos, Winfred Villaluna, and Ethel Ong. 2020. At home with Alexa: a tale of two conversational agents. In *Text, Speech, and Dialogue: 23rd International Conference, TSD 2020, Brno, Czech Republic, September 8–11, 2020, Proceedings 23*. Springer, 495–503.

[36] David L Vannette and Jon A Krosnick. 2014. Answering questions: a comparison of survey satisficing and mindlessness. *The Wiley Blackwell handbook of mindfulness* (2014), 312–327.

[37] Diana Xu, Janet C Read, Emanuela Mazzone, and Martin Brown. 2007. Designing and testing a tangible interface prototype. In *Proceedings of the 6th international conference on Interaction design and children*. 25–28.

[38] Ying Xu and Mark Warschauer. 2019. Young children's reading and learning with conversational agents. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–8.

[39] Bieke Zaman, Vero Vanden Abeele, and Dirk De Grooff. 2013. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction* 1, 2 (2013), 61–70.